

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-305832

(P2000-305832A)

(43) 公開日 平成12年11月2日 (2000.11.2)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テマコード* (参考)
G 0 6 F 12/00	5 3 5	G 0 6 F 12/00	5 3 5 Z 5 B 0 4 5
	5 3 1		5 3 1 R 5 B 0 8 2
9/46	3 6 0	9/46	3 6 0 D 5 B 0 9 8
15/16	6 4 0	15/16	6 4 0 A

審査請求 有 請求項の数 9 O L (全 9 頁)

(21) 出願番号 特願平11-116853

(22) 出願日 平成11年4月23日 (1999.4.23)

(71) 出願人 000164449

九州日本電気ソフトウェア株式会社

福岡市早良区百道浜2丁目4-1 NEC

九州システムセンター

(72) 発明者 大塚 英明

福岡県福岡市早良区百道浜2-4-1 九

州日本電気ソフトウェア株式会社内

(74) 代理人 100082935

弁理士 京本 直樹 (外2名)

Fターム(参考) 5B045 EE03 EE19 GG01

5B082 DA02 DC06 DE06 FA17

5B098 AA10 DD01 GB05 GD16 GD21

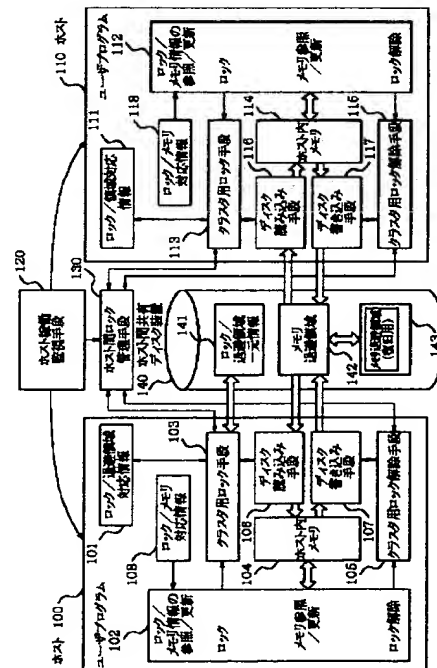
GD22

(54) 【発明の名称】 複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法

(57) 【要約】

【課題】複数ホストから構成されるクラスタシステムにおけるメモリの共有を行う場合、ユーザプログラムの改裝が必要。

【解決手段】ホスト100はユーザプログラム102とクラスタ用ロック手段103とホスト内メモリ104とクラスタ用ロック解除手段105とディスク読み込み手段106とディスク書き込み手段107とロック/メモリ対応情報108とロック/退避領域対応情報101を有しホスト110も同機能があり複数ホスト間の共有機能としてホスト稼動監視手段120とホスト間ロック管理手段130とホスト間共有ディスク装置140がありクラスタ用ロック手段とクラスタ用ロック解除手段実行時に自動的にホスト内メモリを更新することにより1ホスト内で動作していた複数プロセスからメモリを共有するユーザプログラムを改裝なしで複数ホストから構成されるクラスタシステムで実行可能にする。



1

## 【特許請求の範囲】

【請求項1】 複数プロセスからメモリを共有するユーザプログラムと、クラスタ用ロック手段と、ホスト内メモリと、クラスタ用ロック解除手段と、ディスク読み込み手段と、ディスク書き込み手段と、ロック／メモリ対応情報と、ロック／退避領域対応情報とを備えたホストと、複数ホスト間にホスト稼動監視手段と、ホスト間ロック管理手段と、ホスト間共有ディスク装置とを備え、クラスタ用ロック手段およびクラスタ用ロック解除手段実行時、自動的にホスト内メモリを更新することにより、ホスト内で動作していた複数プロセスからメモリを共有するユーザプログラムを、ユーザプログラムの改裝なしで複数ホストから構成されるクラスタシステムで実行可能にすることを特徴とする複数ホストから構成されるクラスタシステムにおけるメモリの共有装置。

【請求項2】 ホスト稼動監視手段により各ホストの稼動状況を監視し、ホスト間ロック管理手段によりホスト間でロック制御を行い、ホスト間共有ディスク装置はロック識別子とメモリ退避領域を管理するロック／退避領域一元情報とロック識別子毎に存在するメモリ退避領域およびメモリ退避領域（復旧用）から構成し、各ホスト毎に存在するロック／メモリ対応情報はロック識別子とそのロック識別子に対応する共有メモリであるホスト内メモリのメモリアドレスとホスト内メモリのメモリサイズを持ちロック識別子をキーに情報が管理され、各ホスト毎に存在するロック／退避領域対応情報およびホスト間共有ディスク装置上に記録されるロック／退避領域一元情報はロック識別子とそのロック識別子に対応する共有メモリの退避領域であるメモリ退避領域の情報を管理する退避領域情報とメモリ退避領域（復旧用）の情報を管理する退避領域情報（復旧用）を持ちロック識別子をキーに情報が管理され、ホスト間共有ディスク装置のメモリ退避領域はホスト間共有ディスク装置へのレコード書き込み中にホストダウンによりアクセス障害が発生した場合のレコードの正当性のチェックをレコードの先頭部の書き込みチェック領域Aと終端部の書き込みチェック領域Bのデータが正しければその間のメモリ退避領域の値も正しい事が保証されるレコードフォーマットにより行うよう構成されることを特徴とする複数ホストから構成されるクラスタシステムにおけるメモリの共有装置。

【請求項3】 ホスト間共有ディスク装置へのレコード書き込み中にホストダウンによりアクセス障害が発生した場合のレコードの正当性のチェックをレコードの先頭部と終端部のデータが正しければその間の値も正しい事が保証されるレコードフォーマットにより行うことを特徴とする複数ホストから構成されるクラスタシステムにおけるメモリの共有装置。

【請求項4】 ロック／メモリ対応情報は、ホスト内で最初にユーザプログラムが実行された時に、ホスト内メ

2

モリの領域確保やロック識別子の取得と同期して作成され、ロック／退避領域対応情報は、クラスタ用ロック手段実行時に参照され、ユーザプログラムからロックが要求されたロック識別子に対応する情報が登録されていなければ、ロック／退避領域一元情報を参照し、登録されていなければ、新たにメモリ退避領域およびメモリ退避領域（復旧用）をそのロック識別子用に確保し、その領域をアクセスするための情報をロック／退避領域一元情報とロック／退避領域対応情報に登録し、メモリ退避領域およびメモリ退避領域（復旧用）は領域の確保時に初期化し、クラスタ用ロック手段は、ホスト間ロック管理手段に対して、ロックを要求しロック成功後に、ロック／退避領域対応情報とロック／メモリ対応情報を参照し、ディスク読み込み手段を利用して、メモリ退避領域からホスト内メモリにデータを読み込み、ユーザプログラムはロック後に、ホスト内メモリの参照および更新を行い、クラスタ用ロック解除手段は、ロック／退避領域対応情報とロック／メモリ対応情報を参照し、ディスク書き込み手段を利用して、メモリ退避領域にホスト内メモリのデータを書き込んだ後、ホスト間ロック管理手段に対して、ロック解除を要求し、ホスト間ロック管理手段は、各ホストからのロック要求を管理するとともに、あるホストがダウンした場合、ホスト稼動監視手段から通知を受け、ダウンしたホストが取得していたロックを解除することを特徴とする複数ホストから構成されるクラスタシステムにおけるメモリの共有方法。

【請求項5】 ユーザプログラムがホスト内で最初に実行された時、ホスト内メモリの領域確保やロック識別子の取得を行い、ロック／メモリ対応情報に登録し、ユーザプログラムが実行されホスト内メモリを参照または更新時にロックを行うために、クラスタ用ロック手段が実行され、クラスタ用ロック手段は、ホスト間ロック管理手段と連携し、クラスタシステム全体のロックが成功した場合、ユーザプログラムが要求したロックのモードにより、参照ロックの場合、参照ロックは、共有メモリを排他的に参照するため、ユーザプログラムからのロック要求時にクラスタ用ロック手段はロック／メモリ対応情報とロック／退避領域対応情報を参照し、ロック識別子に対応するメモリアドレスおよびメモリサイズ、退避領域情報、退避領域情報（復旧用）を取得し、ディスク読み込み手段を使用してホスト間共有ディスク装置のメモリ退避領域から、ホスト内メモリにデータを読み込み、メモリ退避領域のレコードの先頭部の書き込みチェック領域とレコードの終端部の書き込みチェック領域をチェックし、ホストダウンによりデータが不正な場合は、メモリ退避領域（復旧用）から、ホスト内メモリにデータを読み込み、メモリ退避領域（復旧用）の情報をメモリ退避領域に書き込んだ後、ユーザプログラムに制御を戻し、ユーザプログラムから参照ロックのロック解除要求時に、クラスタ用ロック解除手段はホスト間ロック管理

手段と連携し、ロック解除を行い、更新ロックの場合、更新ロックは、共有メモリを排他的に更新するため、ユーザプログラムからのロック要求時にクラスタ用ロック手段はロック／メモリ対応情報とロック／退避領域対応情報を参照し、ロック識別子に対応するメモリアドレスおよびメモリサイズ、退避領域情報、退避領域情報（復旧用）を取得し、ディスク読み込み手段を使用してホスト間共有ディスク装置のメモリ退避領域から、ホスト内メモリにデータを読み込み、メモリ退避領域のレコードの先頭部の書き込みチェック領域とレコードの終端部の書き込みチェック領域をチェックし、ホストダウンによりデータが不正な場合は、メモリ退避領域（復旧用）から、ホスト内メモリにデータを読み込み、メモリ退避領域（復旧用）の情報をメモリ退避領域に書き込んだ後、ユーザプログラムに制御を戻し、メモリ退避領域のデータが正しい場合はメモリ退避領域の内容をディスク書き込み手段を使用してメモリ退避領域（復旧用）に書き込み、ユーザプログラムから更新ロックのロック解除要求時に、クラスタ用ロック解除手段はディスク書き込み手段を使用してホスト内メモリをホスト間共有ディスク装置のメモリ退避領域に書き込み、ホスト間ロック管理手段と連携し、ロック解除を行うことを特徴とする複数ホストから構成されるクラスタシステムにおけるメモリの共有方法。

【請求項6】 あるホストでロック中にホストダウンが発生した場合、ホスト内メモリの領域確保やロック識別子の取得を行い、ロック／メモリ対応情報に登録し、ユーザプログラムが実行されホスト内メモリを参照または更新時にロックを行うために、クラスタ用ロック手段が実行され、クラスタ用ロック手段は、ホスト間ロック管理手段と連携し、クラスタシステム全体のロックが成功した場合、ユーザプログラムが要求したロックのモードにより、参照ロックの場合、参照ロックは、共有メモリを排他的に参照するため、メモリに対する復旧処理は不要で、ロックに関しては、ホストでホストダウンが発生した場合、ホスト稼働監視手段は、ホストのダウンを認識すると、ホスト間ロック管理手段に通知し、ホスト間ロック管理手段はホストからのロックをすべて解除し、ロック解除により動作可能となったユーザプログラムからのロック要求時にクラスタ用ロック手段はロック／メモリ対応情報とロック／退避領域対応情報を参照し、ロック識別子に対応するメモリアドレスおよびメモリサイズ、退避領域情報、退避領域情報（復旧用）を取得し、ディスク読み込み手段を使用してホスト間共有ディスク装置のメモリ退避領域から、ホスト内メモリにデータを読み込み、メモリ退避領域のレコードの先頭部の書き込みチェック領域とレコードの終端部の書き込みチェック領域をチェックし、ホストダウンによりデータが不正な場合は、メモリ退避領域（復旧用）から、ホスト内メモリにデータを読み込み、メモリ退避領域（復旧用）の情

報をメモリ退避領域に書き込んだ後、ユーザプログラムに制御を戻し、ユーザプログラムから参照ロックのロック解除要求時に、クラスタ用ロック解除手段はホスト間ロック管理手段と連携し、ロック解除を行い、更新ロックの場合更新ロックは、共有メモリを排他的に更新するため、メモリに対する復旧処理は不要で、ロック解除により動作可能となったユーザプログラムからのロック要求時にクラスタ用ロック手段はロック／メモリ対応情報とロック／退避領域対応情報を参照し、ロック識別子に対応するメモリアドレスおよびメモリサイズ、退避領域情報、退避領域情報（復旧用）を取得し、ディスク読み込み手段を使用してホスト間共有ディスク装置のメモリ退避領域から、ホスト内メモリにデータを読み込み、メモリ退避領域のレコードの先頭部の書き込みチェック領域とレコードの終端部の書き込みチェック領域をチェックし、ホストダウンによりデータが不正な場合は、メモリ退避領域（復旧用）から、ホスト内メモリにデータを読み込み、メモリ退避領域（復旧用）の情報をメモリ退避領域に書き込んだ後、ユーザプログラムに制御を戻し、メモリ退避領域のデータが正しい場合はメモリ退避領域の内容をディスク書き込み手段を使用してメモリ退避領域（復旧用）に書き込み、ユーザプログラムから更新ロックのロック解除要求時に、クラスタ用ロック解除手段はディスク書き込み手段を使用してホスト内メモリをホスト間共有ディスク装置のメモリ退避領域に書き込み、ホスト間ロック管理手段と連携し、ロック解除を行うことを特徴とする複数ホストから構成されるクラスタシステムにおけるメモリの共有方法。

【請求項7】 各ホストのユーザプログラムが全て参照ロックの場合、ホストのユーザプログラムは、ロック／対応情報を参照し、共有メモリを参照するためのロック（参照ロック）を行い、ホスト間共有ディスク装置中のメモリ退避領域からホスト内メモリにデータを読み込み、他のホストのユーザプログラムは自己のロック／対応情報を参照し、共有メモリを参照するためのロック（参照ロック）を行い、参照ロックであるためロックに成功し、ホスト間共有ディスク装置中のメモリ退避領域から自己のホスト内メモリにデータを読み込むことを特徴とする複数ホストから構成されるクラスタシステムにおけるメモリの共有方法。

【請求項8】 参照ロックのユーザプログラムと更新ロックのユーザプログラムの場合、参照ロックのホストのユーザプログラムは、ロック／対応情報を参照し、共有メモリを参照するためのロック（参照ロック）を行い、ホスト間共有ディスク装置中のメモリ退避領域からホスト内メモリにデータを読み込み、更新ロックのホストのユーザプログラムは自己のロック／対応情報を参照し、共有メモリを更新するためのロック（更新ロック）を行うことによりユーザプログラムで参照ロック中であるためロック待ち状態となり、ユーザプログラムでロックが

解除された後でロックが成功し、ホスト間共有ディスク装置中のメモリ退避領域からホスト内メモリにデータが読み込まれ、更新ロックのユーザプログラムでロック解除時に、自己のホスト内メモリは、ホスト間共有ディスク装置中のメモリ退避領域に書き込まれることを特徴とする複数ホストから構成されるクラスタシステムにおけるメモリの共有方法。

【請求項 9】 各ホストのユーザプログラムが全て更新ロックの場合、ホストのユーザプログラムは、ロック／対応情報を参照し、共有メモリを更新するためのロック（更新ロック）を行い、ホスト間共有ディスク装置中のメモリ退避領域からホスト内メモリにデータが読み込み、他のホストのユーザプログラムは自己のロック／対応情報を参照し、共有メモリを更新するためのロック（更新ロック）を行うことによりユーザプログラムで更新ロック中であるためロック待ち状態となり、ユーザプログラムがロック解除時に、ホスト内メモリは、ホスト間共有ディスク装置中のメモリ退避領域に書き込まれ、ロックが解除された後で他のユーザプログラムでロックが成功し、ホスト間共有ディスク装置中のメモリ退避領域からホスト内メモリにデータが読み込まれ、他のユーザプログラムでロック解除時に、自己のホスト内メモリは、ホスト間共有ディスク装置中のメモリ退避領域に書き込まれることを特徴とする複数ホストから構成されるクラスタシステムにおけるメモリの共有方法。

#### 【発明の詳細な説明】

##### 【0001】

【発明の属する技術分野】本発明は複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法に関し、特に 1 ホスト内で動作実績がある複数プロセスからメモリを共有するユーザプログラムを、ユーザプログラムの改裝なしで複数ホストから構成されるクラスタシステムで実行可能とする複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法に関する。

##### 【0002】

【従来の技術】従来のメモリ共有型のマルチプロセスシステムの場合のメモリ共有方式は、特開平 08-166933 号公報に記載されている。この従来のメモリ共有方式は、各ホストに存在するホスト間共有データ領域と、ホスト間共有データ領域所在管理手段と、データ送受信手段から構成されている。

【0003】このような構成を有する従来のメモリ共有方式はつぎのように動作する。

【0004】すなわち、プログラムがホスト間共有データ領域をアクセスすると、ホスト間共有データ領域管理手段が作動し、他ホストと通信を行い、ホスト間共有データ領域を補正している。

##### 【0005】

【発明が解決しようとする課題】上述した従来のメモリ

共有方式は、第 1 の問題点は、すでに 1 ホスト上で動作していた既存ユーザプログラムの改造が必要であるということである。

【0006】その理由は、ホスト間共有データ領域をアクセスするように改造しなければならないためである。

【0007】第 2 の問題点は、処理性能が遅いということである。

【0008】その理由は、他ホストとの通信が必要であり、ホスト台数に比例して通信量が増加するためである。また、メモリを排他的にアクセスしたい場合でも、他ホストへの確認が必要になる点である。

##### 【0009】

【課題を解決するための手段】本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法は、複数プロセスからメモリを共有するユーザプログラムと、クラスタ用ロック手段と、ホスト内メモリと、クラスタ用ロック解除手段と、ディスク読み込み手段と、ディスク書き込み手段と、ロック／メモリ対応情報と、ロック／退避領域対応情報とを備えたホストと、複数ホスト間にホスト稼働監視手段と、ホスト間ロック管理手段と、ホスト間共有ディスク装置とを備え、ホスト稼働監視手段により各ホストの稼働状況を監視し、ホスト間ロック管理手段によりホスト間でロック制御を行い、ホスト間共有ディスク装置はロック識別子とメモリ退避領域を管理するロック／退避領域一元情報とロック識別子毎に存在するメモリ退避領域およびメモリ退避領域（復旧用）から構成し、各ホスト毎に存在するロック／メモリ対応情報はロック識別子とそのロック識別子に対応する共有メモリであるホスト内メモリのメモリアドレスとホスト内メモリのメモリサイズを持ちロック識別子をキーに情報が管理され、各ホスト毎に存在するロック／退避領域対応情報およびホスト間共有ディスク装置上に記録されるロック／退避領域一元情報はロック識別子とそのロック識別子に対応する共有メモリの退避領域であるメモリ退避領域の情報を管理する退避領域情報とメモリ退避領域（復旧用）の情報を管理する退避領域情報（復旧用）を持ちロック識別子をキーに情報が管理され、ホスト間共有ディスク装置のメモリ退避領域はホスト間共有ディスク装置へのレコード書き込み中にホストダウンによりアクセス障害が発生した場合のレコードの正当性のチェックをレコードの先頭部の書き込みチェック領域 A と終端部の書き込みチェック領域 B のデータが正しければその間のメモリ退避領域の値も正しい事が保証されるレコードフォーマットにより行うよう構成され、ロック／メモリ対応情報は、ホスト内で最初にユーザプログラムが実行された時に、ホスト内メモリの領域確保やロック識別子の取得と同期して作成され、ロック／退避領域対応情報は、クラスタ用ロック手段実行時に参照され、ユーザプログラムからロックが要求されたロック識別子に対応する情報が登録されていなければ、ロ

ック／退避領域一元情報を参照し、登録されていなければ、新たにメモリ退避領域およびメモリ退避領域（復旧用）をそのロック識別子用に確保し、その領域をアクセスするための情報をロック／退避領域一元情報とロック／退避領域対応情報に登録し、メモリ退避領域およびメモリ退避領域（復旧用）は領域の確保時に初期化し、クラスタ用ロック手段は、ホスト間ロック管理手段に対して、ロックを要求しロック成功後に、ロック／退避領域対応情報とロック／メモリ対応情報を参照し、ディスク読み込み手段を利用して、メモリ退避領域からホスト内メモリにデータを読み込み、ユーザプログラムはロック後に、ホスト内メモリの参照および更新を行い、クラスタ用ロック解除手段は、ロック／退避領域対応情報とロック／メモリ対応情報を参照し、ディスク書き込み手段を利用して、メモリ退避領域にホスト内メモリのデータを書き込んだ後、ホスト間ロック管理手段に対して、ロック解除を要求し、ホスト間ロック管理手段は、各ホストからのロック要求を管理するとともに、あるホストがダウンした場合、ホスト稼働監視手段から通知を受け、ダウンしたホストが取得していたロックを解除し、クラスタ用ロック手段およびクラスタ用ロック解除手段実行時、自動的にホスト内メモリを更新することにより、ホスト内で動作していた複数プロセスからメモリを共有するユーザプログラムを、ユーザプログラムの改装なしで複数ホストから構成されるクラスタシステムで実行可能にするように構成されている。

#### 【0010】

【発明の実施の形態】本発明は、1ホスト内で動作実績がある複数プロセスからメモリを共有するユーザプログラムを、ユーザプログラムの改装なしで複数ホストから構成されるクラスタシステムで実行可能とするための方式である。ここで言うクラスタシステムとは、同一構成のシステムを複数台のホストに実装し、各ホストの負荷状況に応じて、処理を自動的に負荷分散するシステムのことである。

【0011】次に、本発明の実施の形態について図面を参照して説明する。

【0012】図1は本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法の一実施の形態を示すブロック図である。

【0013】図1を参照すると、本発明の実施の形態はホスト100と、ホスト110と、各ホストの稼働を監視するホスト稼働監視手段120と、ホスト間のロック制御を行うホスト間ロック管理手段130と、ホスト間で共有するホスト間共有ディスク装置140から構成されている。

【0014】クラスタシステムを構成するホスト100には、ユーザプログラム102と、ロック機能のユーザインタフェースを提供するクラスタ用ロック手段103と、ユーザプログラム102が他プログラムと共有する

メモリ領域であるホスト内メモリ104と、ロック解除機能のユーザインタフェースを提供するクラスタ用ロック解除手段105と、ホスト間共有ディスク装置140からデータをホスト内メモリ104に読み込むディスク読み込み手段106と、ホスト内メモリ104からホスト間共有ディスク装置140にデータを書き込むディスク書き込み手段107と、ロック識別子と共有メモリ領域の対応を管理するロック／メモリ対応情報108と、ロック識別子とメモリの退避領域の対応を管理するロック／退避領域対応情報101から構成される。

【0015】ホスト110は、ホスト100と同じ構成であり、図1ではホスト2台の構成となっているが、ホスト台数は自由に増やすことができる。

【0016】ホスト稼働監視手段120は、各ホストの稼働状況を監視するために使用される。

【0017】ホスト間ロック管理手段130は、ホスト間でロック制御を行うために使用される。

【0018】ホスト間共有ディスク装置140は、ロック識別子やメモリ退避領域を管理するロック／退避領域一元情報141と、ロック識別子毎に存在するメモリ退避領域142およびメモリ退避領域（復旧用）143から構成される。

【0019】各ホスト毎に存在するロック／メモリ対応情報108には、図3の本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法におけるロック識別子と、そのロック識別子に対応する共有メモリであるホスト内メモリのメモリアドレスと、メモリサイズとの対応情報の構造を示す図に示すように、ロック識別子301と、そのロック識別子に対応する共有メモリであるホスト内メモリ104のメモリアドレス302と、ホスト内メモリ104のメモリサイズ303を持ち、ロック識別子301をキーに情報が管理される。

【0020】各ホスト毎に存在するロック／退避領域対応情報101およびホスト間共有ディスク装置140上に記録されるロック／退避領域一元情報141には、図4の本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法におけるロック識別子と、そのロック識別子に対応する共有メモリの退避領域情報と、退避領域情報（復旧用）との対応情報の構造を示す図に示すように、ロック識別子401と、そのロック識別子に対応する共有メモリの退避領域であるメモリ退避領域142の情報を管理する退避領域情報402と、メモリ退避領域（復旧用）143の情報を管理する退避領域情報（復旧用）403を持ち、ロック識別子401をキーに情報が管理される。

【0021】ホスト間共有ディスク装置140のメモリ退避領域142は、図5の本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法におけるホスト間共有ディスク装置のメモリ退避領域

9

の書き込みチェック領域と、メモリ退避領域と、書き込みチェック領域との対応情報の構造を示す図に示すように、書き込みチェック領域A501と、メモリ退避領域502と、書き込みチェック領域B503から構成される。

【0022】これらの手段はそれぞれつぎのように動作する。なお、説明はホスト100およびホスト間で共有する手段等について行うが、ホスト110の各手段も同じ動作をする。

【0023】ロック／メモリ対応情報108は、ホスト100内で最初にユーザプログラム102が実行された時などに、ホスト内メモリ104の領域確保やロック識別子の取得と同期して作成される。

【0024】ロック／退避領域対応情報101は、クラスタ用ロック手段103実行時に参照され、ユーザプログラム102からロックが要求されたロック識別子に対応する情報が登録されていないければ、ロック／退避領域一元情報141を参照し、そこにも登録されていないければ、新たにメモリ退避領域142およびメモリ退避領域（復旧用）143をそのロック識別子用に確保し、その領域をアクセスするための情報をロック／退避領域一元情報141とロック／退避領域対応情報101に登録する。

【0025】メモリ退避領域142およびメモリ退避領域（復旧用）143は領域の確保時に初期化する。

【0026】クラスタ用ロック手段103は、図2の本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法におけるクラスタシステムを構成する1ホストであるメモリ共有型のマルチプロセッサシステムを示すブロック図に示すメモリ共有型マルチプロセッサシステムであるホスト200上の、ロック手段205と互換性があるユーザインタフェースを持ち、ホスト間ロック管理手段130に対して、ロックを要求しロック成功後に、ロック／退避領域対応情報101とロック／メモリ対応情報108を参照し、ディスク読み込み手段106を利用して、メモリ退避領域142からホスト内メモリ104にデータを読み込む。

【0027】ユーザプログラム102はロック後に、ホスト内メモリ104の参照および更新を行う。

【0028】ユーザプログラム102は、図2で示されるメモリ共有型のマルチプロセッサシステム上で動作するユーザプログラム202およびユーザプログラム203と同一のプログラムである。

【0029】クラスタ用ロック解除手段105は、図2で示すメモリ共有型マルチプロセッサシステムであるホスト200上の、ロック解除手段207と互換性があるユーザインタフェースを持ち、ロック／退避領域対応情報101とロック／メモリ対応情報108を参照し、ディスク書き込み手段107を利用して、メモリ退避領域142にホスト内メモリ104のデータを書き込んだ後

10

で、ホスト間ロック管理手段130に対して、ロック解除を要求する。

【0030】ホスト間ロック管理手段130は、各ホストからのロック要求を管理し、表1のロック制御例に示すようなサービスを提供する。

【0031】

【表1】

	参照ロック	更新ロック
参照ロック	同時実行可能	同時実行不可能
更新ロック	同時実行不可能	同時実行不可能

【0032】また、ホスト間ロック管理手段130は、あるホストがダウンした場合は、ホスト稼動監視手段120から通知を受け、ダウンしたホストが取得していたロックを解除する。

【0033】ホスト間共有ディスク装置140の特性として、レコード書き込み中にホストダウン等によりアクセス障害が発生した場合、レコードの先頭部と終端部のデータが正しければその間の値も正しい事が保証される場合のレコードフォーマットを図5に示している。実際に使用するディスク装置の特性にあったレコードフォーマットに変更する必要がある。

【0034】メモリ退避領域（復旧用）143は、ホストダウン等によりメモリ退避領域142へのデータ書き込みに失敗した場合に、メモリ退避領域142を復旧するために使用される。

【0035】次に、本発明の実施の形態の動作について、図1～図5および表1を参照して詳細に説明する。

【0036】まず初めに、図2によりクラスタシステムを構成する1ホストであるメモリ共有型のマルチプロセッサシステム上における従来動作を説明する。

【0037】ホスト200で動作するユーザプログラム202または203は、ホスト内で最初に実行された時などにホスト内メモリ206の領域確保やロック識別子の取得を行い、ロック／メモリ対応情報204に登録する。

【0038】ロック／メモリ対応情報204は、ユーザプログラム202または203から参照され、ホスト内メモリ206アクセス時のロックやメモリアドレス等の取得に使用される。

【0039】ユーザプログラム202または203からロック要求が行われると、ロック手段205は、ホスト内ロック管理手段201を使用してロックを行う。ホスト内ロック管理手段201は表1に示すようなサービスを提供する。

【0040】ユーザプログラム202または203は、ロック後にホスト内メモリ206のアクセスを行う。

【0041】ユーザプログラム202または203からロック解除要求が行われると、ロック解除手段207は、ホスト内ロック管理手段201を使用してロック解除を行う。

【0042】次にクラスタシステムの場合の動作について説明する。

【0043】ユーザプログラム102はホスト100内で最初に行われた時などにホスト内メモリ108の領域確保やロック識別子の取得を行い、ロック／メモリ対応情報108に登録する。

【0044】ユーザプログラム102が実行されホスト内メモリ104を参照または更新時にロックを行うために、クラスタ用ロック手段103が実行される。クラスタ用ロック手段103は、ホスト間ロック管理手段130と連携し、クラスタシステム全体のロックが成功した場合、ユーザプログラム102が要求したロックのモードにより、次の動作を行う。

【0045】(11) 参照ロックの場合

参照ロックは、共有メモリを排他的に参照することが目的であるため、次のように動作する。

【0046】ユーザプログラム102からのロック要求時にクラスタ用ロック手段103はロック／メモリ対応情報108とロック／退避領域対応情報101を参照し、図3および図4で示すロック識別子に対応するメモリアドレス302およびメモリサイズ303、退避領域情報402、退避領域情報(復旧用)403を取得し、ディスク読み込み手段106を使用してホスト間共有ディスク装置140のメモリ退避領域142から、ホスト内メモリ104にデータを読み込む。この時、メモリ退避領域142中の図5に示す書き込みチェック領域A501と書き込みチェック領域B503をチェックし、ホストダウン等によりデータが不正な場合は、メモリ退避領域(復旧用)143から、ホスト内メモリ104にデータを読み込み、メモリ退避領域(復旧用)143の情報をメモリ退避領域142に書き込んだ後でユーザプログラム102に制御を戻す。

【0047】ユーザプログラム102から参照ロックのロック解除要求時に、クラスタ用ロック解除手段105はホスト間ロック管理手段130と連携し、ロック解除を行う。

【0048】(12) 更新ロックの場合

更新ロックは、共有メモリを排他的に更新することが目的であるため、次のように動作する。

【0049】ユーザプログラム102からのロック要求時にクラスタ用ロック手段103はロック／メモリ対応情報108とロック／退避領域対応情報101を参照し、図3および図4で示すロック識別子に対応するメモリアドレス302およびメモリサイズ303、退避領域情報402、退避領域情報(復旧用)403を取得し、ディスク読み込み手段106を使用してホスト間共有デ

ィスク装置140のメモリ退避領域142から、ホスト内メモリ104にデータを読み込む。この時、メモリ退避領域142中の図5に示す書き込みチェック領域A501と書き込みチェック領域B503をチェックし、ホストダウン等によりデータが不正な場合は、メモリ退避領域(復旧用)143から、ホスト内メモリ104にデータを読み込み、メモリ退避領域(復旧用)143の情報をメモリ退避領域142に書き込んだ後でユーザプログラム102に制御を戻す。メモリ退避領域142のデータが正しい場合はメモリ退避領域142の内容をディスク書き込み手段107を使用してメモリ退避領域(復旧用)143に書き込む。

【0050】ユーザプログラム102から更新ロックのロック解除要求時に、クラスタ用ロック解除手段105はディスク書き込み手段107を使用してホスト内メモリ104をホスト間共有ディスク装置140のメモリ退避領域142に書き込み、ホスト間ロック管理手段130と連携し、ロック解除を行う。

【0051】次に、あるホストでロック中にホストダウンが発生した場合の動作について説明する。

【0052】(21) 参照ロックの場合

参照ロックは、共有メモリを排他的に参照することが目的であるため、メモリに対する復旧処理は不要である。

【0053】ロックに関しては、ホスト110でホストダウンが発生した場合を例に説明する。ホスト稼働監視手段120は、ホスト110のダウンを認識すると、ホスト間ロック管理手段130に通知する。ホスト間ロック管理手段130はホスト110からのロックをすべて解除する。ロック解除により動作可能となったホスト100のユーザプログラムは、前述の処理を行う。

【0054】(22) 更新ロックの場合

更新ロックは、共有メモリを排他的に更新することが目的であるが、メモリに対する復旧処理は不要である。

【0055】ロックの解除については、参照ロック時と同じしくみで動作する。ロック解除により動作可能となったホスト100のユーザプログラムは、前述の処理を行う。ホスト間共有ディスク装置140のメモリ退避領域142にホスト内メモリ104のデータを書き込み中に障害が発生した場合でも、ロック時にデータのチェックが行われるため前回のロック解除成功時の正しいデータを参照することができる。

【0056】次に、具体的な例を用いて説明する。

【0057】(31) ユーザプログラム102とユーザプログラム112がともに参照ロックの場合

ホスト100のユーザプログラム102は、ロック／対応情報108を参照し、共有メモリを参照するためのロック(参照ロック)を行う。この時、ホスト間共有ディスク装置140中のメモリ退避領域142からホスト内メモリ104にデータが読み込まれる。次に、ホスト110のユーザプログラム112がロック／対応情報11



13

8を参照し、共有メモリを参照するためのロック（参照ロック）を行う。参照ロックであるためロックに成功し、ユーザプログラム102と同様にホスト内メモリ104にデータが読み込まれる。

【0058】（32）ユーザプログラム102が参照ロックでユーザプログラム112が更新ロックの場合ホスト100のユーザプログラム102は、ロック／対応情報108を参照し、共有メモリを参照するためのロック（参照ロック）を行う。この時、ホスト間共有ディスク装置140中のメモリ退避領域142からホスト内メモリ104にデータが読み込まれる。次に、ホスト110のユーザプログラム112がロック／対応情報118を参照し、共有メモリを更新するためのロック（更新ロック）を行うとユーザプログラム102で参照ロック中であるためロック待ち状態となる。ユーザプログラム102でロックが解除された後でロックが成功し、ホスト間共有ディスク装置140中のメモリ退避領域142からホスト内メモリ104にデータが読み込まれる。ユーザプログラム112でロック解除時に、ホスト内メモリ114は、ホスト間共有ディスク装置140中のメモリ退避領域142に書き込まれる。

【0059】（33）ユーザプログラム102とユーザプログラム112がともに更新ロックの場合ホスト100のユーザプログラム102は、ロック／対応情報108を参照し、共有メモリを更新するためのロック（更新ロック）を行う。この時、ホスト間共有ディスク装置140中のメモリ退避領域142からホスト内メモリ104にデータが読み込まれる。次に、ホスト110のユーザプログラム112がロック／対応情報118を参照し、共有メモリを更新するためのロック（更新ロック）を行うとユーザプログラム102で更新ロック中であるためロック待ち状態となる。ユーザプログラム102がロック解除時に、ホスト内メモリ104は、ホスト間共有ディスク装置140中のメモリ退避領域142に書き込まれ、ロックが解除された後でユーザプログラム112でロックが成功し、ホスト間共有ディスク装置140中のメモリ退避領域142からホスト内メモリ104にデータが読み込まれる。ユーザプログラム112でロック解除時に、ホスト内メモリ114は、ホスト間共有ディスク装置140中のメモリ退避領域142に書き込まれる。

【0060】

【発明の効果】以上説明したように、本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法は、第1の効果は、1ホストで構築していたシステム上で動作している多数のユーザプログラムを改造なしで複数ホストで構成されるクラスタシステムへ移行できることにある。

【0061】その理由は、もし、ロック／アンロック機能と連動したメモリ共有ができない場合、1ホスト内で

14

しかユーザプログラムを実行できないが、本発明を利用することにより、ユーザプログラムを改造なしで複数ホストで実行できるからである。また、本発明を使用せず、ユーザプログラムをクラスタシステムに合った処理ロジックに改造する方法もあるが、その場合、多大な修正工数を要す。

【図面の簡単な説明】

【図1】本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法の一実施の形態を示すブロック図である。

【図2】本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法におけるクラスタシステムを構成する1ホストであるメモリ共有型のマルチプロセスシステムを示すブロック図である。

【図3】本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法におけるロック識別子と、そのロック識別子に対応する共有メモリであるホスト内メモリのメモリアドレスと、メモリサイズとの対応情報の構造を示す図である。

【図4】本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法におけるロック識別子と、そのロック識別子に対応する共有メモリの退避領域情報と、退避領域情報（復旧用）との対応情報の構造を示す図である。

【図5】本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法におけるホスト間共有ディスク装置のメモリ退避領域の書き込みチェック領域と、メモリ退避領域と、書き込みチェック領域との対応情報の構造を示す図である。

【符号の説明】

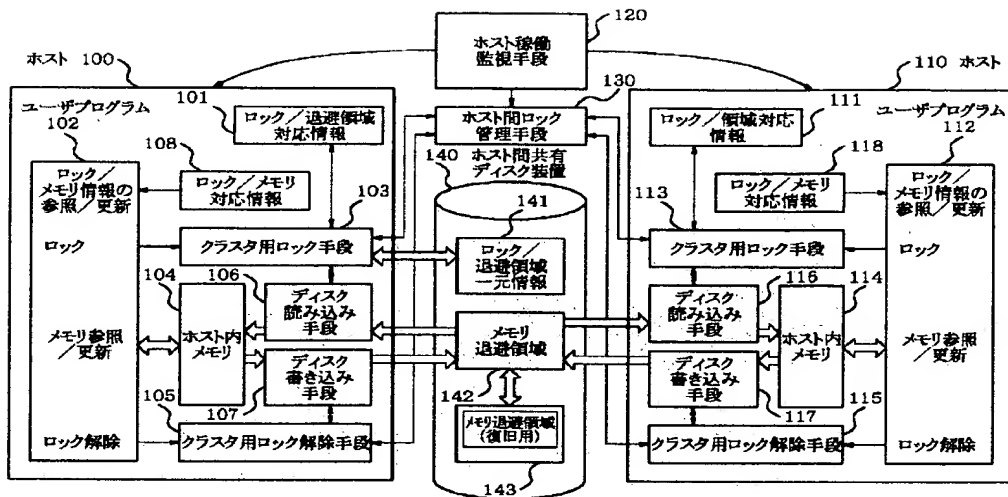
100, 110	ホスト
101	ロック／退避領域対応情報
102, 112	ユーザプログラム
103, 113	クラスタ用ロック手段
104, 114	ホスト内メモリ
105, 115	クラスタ用ロック解除手段
106, 116	ディスク読み込み手段
107, 117	ディスク書き込み手段
108, 118	ロック／メモリ対応情報
120	ホスト稼動監視手段
130	ホスト間ロック管理手段
140	ホスト間共有ディスク装置
141	ロック／退避領域一元情報
142	メモリ退避領域
143	メモリ退避領域（復旧用）
301, 401	ロック識別子
302	メモリアドレス
303	メモリサイズ
402	退避領域情報
403	退避領域情報（復旧用）



15  
501 書き込みチェック領域A  
502 メモリ回避領域

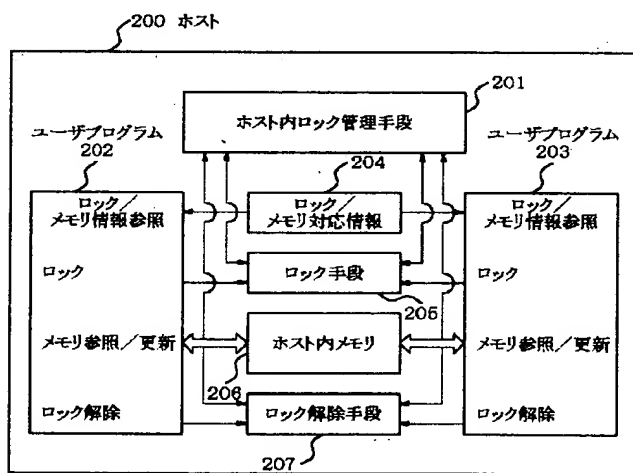
16  
\*503 書き込みチェック領域B

【図1】



【図2】

【図3】



【図4】

301 ロック識別子	302 メモリアドレス	303 メモリサイズ
.....	.....	.....
~~~~~	~~~~~	~~~~~

【図5】

401 ロック識別子	402 回避領域情報	403 回避領域情報(復旧用)
.....	.....	.....
~~~~~	~~~~~	~~~~~

501 書き込みチェック領域	502 メモリ回避領域	503 書き込みチェック領域
~~~~~	~~~~~	~~~~~